# Fundamentals of Python Programming

Ruwen Qin
Stony Brook University

August 22, 2025

For a data science and machine learning course, a solid foundation in Python is essential. First, this module will guide you to set up the Python coding environment. Then, it will introduce Numpy for array and vectorized computation and Pandas comprising data structures and data manipulation tools. After that, students will learn data file reading and writing, data cleaning, and preparation.

## 1 Python Programming Environment

### 1.1 Google Colab

Google Colab (colab.google.com) is a cloud-based service that hosts Jupyter notebooks and provides free access to computing resources, including GPUs and TPUs, without any setup. Students who are concerned about their personal computer's capability can consider using Google Colab for this course.

- Install Google Drive for Desktop (https://ipv4.google.com/intl/en_zm/drive/download/) so that Google Drive is shown as a folder in your computer. With that setup, accessing files in your Google Drive is just like access files in your local computer, which is better than using a browser. Sign in your Google Drive using your SBU account.
- Log in Colab using your Stony Brook account.
- Mount your Google Drive to Colab, which allows you to read files from, and save files to, Google Drive.
- Create a folder named "Colab Notebooks" in your Stony Brook Google Drive, dedicated to this course. Within that folder, create a subfolder named "Data" to save data files for this course.
- Make the "Colab Notebooks" folder as the current working directory.

### 1.2 Jupyter Notebook

Jupyter Notebook is an open-source application that runs on your computer locally. Students who would like to have complete control over their environment, data, and hardware can choose this option.

- Install Miniconda (https://www.anaconda.com/docs/getting-started/miniconda/install)
- It is recommended to create an environment dedicated to this course. For example, we create an environment called "CIV355" with the version of python to be 3.9.23 by running the following script in Terminal:

```
        conda create -n CIV355 python=3.9.23

```

- Activate this environment:

```
        conda activate CIV355

```

- Install Jupyter notebook within this envirionment:

```
        conda install jupyter notebook

```

- Install other packages such as Numpy and Pandas within this environment:

```
conda install numpy pandas
```

- Install other packages within this environment whenever we need more
- To deactivate the environment, run the following script:

```
conda deactivate
```

When we would like to use Jupyter Notebook for this course, we will always first activate the environment CIV355 from the terminal, and then run the following script:

```
jupyter notebook
```

Then, it launches a local web server on your computer.

It is recommended that you create a folder to save your Jupyter notebooks and data for this course. Jupyter Notebook cannot access online accounts. Therefore, if you would like to read and write files in Google Drive, you will have to install the app - Google Drive for Desktop - to your computer to have a local folder for your Google Drive. Installing the app Google Drive for Desktop and accessing it using your SBU account can be complex for some operating systems. If you wish to save your files in a cloud platform, a simpler solution is One Drive. The easiest solution is that you save your files in a local folder within your personal computer.

## 1.3 Other Python Code Editors and IDE

Feel free to use other code editors (e.g., VSCode) or Integrated Development Environment (IDE) (e.g., PyCharm). We skip the details, and you can choose one for use based on your personal preference.